Cell-Based Metrics Improve the Detection of Gene-Gene Interactions Using Multifactor Dimensionality Reduction

Jonathan M. Fisher[1], Peter Andrews[1], Jeff Kiralis[1], Nicholas A. Sinnott-Armstrong[1], and Jason H. Moore[1,2,3]

[1] Department of Genetics, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA
[2] Department of Community and Family Medicine, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA
[3] Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH 03755, USA
jonathan.m.fisher@dartmouth.edu
http://www.epistasis.org/

Abstract. Multifactor Dimensionality Reduction (MDR) is a widely-used data-mining method for detecting and interpreting epistatic effects that do not display significant main effects. MDR produces a reduced-dimensionality representation of a dataset which classifies multi-locus genotypes into either high- or low-risk groups. The weighted fraction of cases and controls correctly labelled by this classification, the balanced accuracy, is typically used as a metric to select the best or most-fit model. We propose two new metrics for MDR to use in evaluating models, Variance and Fisher, and compare those metrics to two previously-used MDR metrics, Balanced Accuracy and Normalized Mutual Information. We find that the proposed metrics consistently outperform the existing metrics across a variety of scenarios.

Keywords: Multifactor Dimensionality Reduction; Fisher's exact test.

1 Introduction

Epistasis, or gene-gene interaction, is fundamental in gene expression, and figures prominently in the genetics of complex traits such as susceptibility to disease (e.g., [3,13]). Epistasis introduces complexity in the relationship between genotype and phenotype, making patterns in that relationship more difficult to detect. We therefore need tools that enable us to detect epistasis and search for the patterns that might be hidden behind it [20,8,9,19,21,13,15]. In particular, our tools should make use of all of the information available in each dataset.

Multifactor Dimensionality Reduction (MDR) is a non-parametric data-mining tool that can detect epistatic models of gene expression that do not show significant main effects, widely used in the study of genetic traits with or without a component of environmental causation [18,17,6,5,7,10,11,24]. MDR

L. Vanneschi, W.S. Bush, and M. Giacobini (Eds.): EvoBIO 2013, LNCS 7833, pp. 200–211, 2013.

uses a constructive-induction algorithm to label each genotype combination as high-risk or low-risk based on a discrete endpoint such as case–control status, constructing a new variable with two risk levels which pools all high-risk genotypes into one group and all low-risk genotypes into another group[11,10]. That new variable can then be analyzed with a classification method such as naïve Bayes or logistic regression.

For any desired order of interaction N (typically between 2 and 4), MDR iterates over all sets of N loci and constructs a model for each of them. Each individual in the dataset is classified according to which allele it has at each locus of the model, and a case-control table is constructed which counts how many individuals of each allelic combination are cases and how many are controls (Fig. 1B). That case-control table encapsulates the model for that set of loci. MDR then chooses, from all of the models, the case-control table that scores highest by whatever metric (i.e., measure of model fitness) it uses. Finally, MDR constructs a new variable from that case-control table, labeling each combination of genotype values as low-risk if the corresponding cell in the case-control table has a ratio of cases to controls below a pre-chosen threshold, and high-risk otherwise.

In standard MDR, the metric used is balanced accuracy [24], defined as the mean of sensitivity and specificity:

$$\frac{TP/(TP+FN)+TN/(TN+FP)}{2}, \tag{1}$$

where TP is the number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives. The threshold typically used is the ratio of cases to controls in the dataset as a whole, as recommended by Velez et al [24]. Various alternative metrics to balanced accuracy have been investigated, but they have typically focused on evaluating each possible model simply in terms of the contingency table of the number of true positives, true negatives, false negatives, and false positives that the model produces on the given dataset [1].

We propose to improve the metric used by MDR by making full use of the information available in the dataset – looking at the full case-control table of status versus genotype, instead of the summary table of risk-level versus status. Specifically, we propose two new metrics, one based on a Fisher's exact test applied to the case-control table, and one based on the variance of case-control ratios in the table. We evaluate the metrics for their ability to pick out a known signal from noise in simulated datasets across a wide variety of scenarios, and compare their performance to that of standard metrics. We show that the new metrics display equal or greater detection ability across all of the scenarios investigated, with significantly improved ability to detect weaker signals.

2 Methods

2.1 Use of Metrics

The function of MDR is to construct a new attribute by selecting the model which, for a given dataset, best predicts the phenotype from the genotypic information.

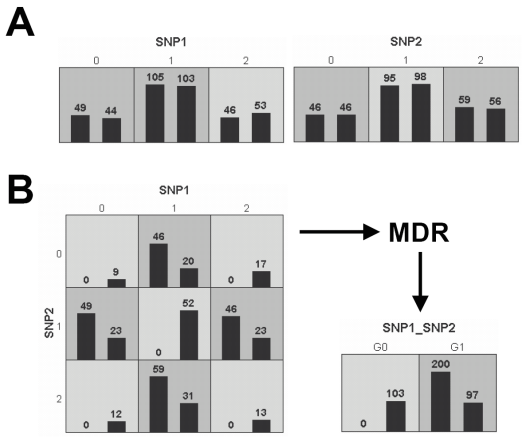

Fig. 1. MDR attribute construction. (A) illustrates the distribution of cases (left bars) and controls (right bars) for each of the three genotypes of SNP1 and SNP2 in an example dataset. The shading of the cells indicates the labeling that has been assigned (using a threshold of T = 1): dark shading indicates "high-risk" and light shading "low risk". (B) illustrates the distribution of cases and controls when the two functional SNPs are considered jointly. A new single attribute is constructed by pooling the high-risk genotype combinations into one group, G1, and the low-risk into another group, G0. Reprinted from Velez et al [24].

Each model is based on a small number of interacting loci, typically between 1 and 4, and yields the case-control table of status versus genotype for those loci over the dataset. MDR works in two phases, first selecting a model within each level of interaction, and then choosing a model from among the levels. Within a given level of interaction, MDR uses a metric to score each case-control table produced from the dataset and then selects the case-control table with the highest score assigned by the metric. We concentrate here on evaluating the metrics used to compare models within a single level of interaction.

The value of each metric on a model is calculated by first constructing the case-control table of the model over the dataset and then applying some formula to the table; the exact nature of the formula is what defines the metric.

2.2 Definitions of Metrics

We propose two new metrics for MDR, Variance and Fisher; we evaluate them by comparing them to two metrics that have been used previously in MDR: Balanced Accuracy [24], which is the standard metric used in MDR, and Normalized Mutual Information, which has been recommended by Bush et al [1]. In each evaluation-run we test the detection ability of each metric, the ability to identify the set of loci that constitute the known signal.

The Variance metric is motivated by the concept of the variance of the case-fractions in a case-control table: for a table with N cells, the Variance metric is defined as the total variance of case-fractions in the table,

$$\sum_{i=1}^{N} p_i(k_i - k)^2 \ , \tag{2}$$

where N is the number of cells in the table, p_i is the fraction of the individuals that lie in the ith cell (which is the sample approximation to the probability that an individual chosen at random lies in the ith cell), k_i is the fraction of individuals in the ith cell that have the condition or trait (i.e., the case-fraction in the cell), and k is the fraction of individuals that have the condition or trait in the dataset as a whole. If the the ith cell cell is empty, we define the ith term of the sum to be 0.

The Fisher metric uses a Fisher's exact test to measure how unusual each cell of a case-control table is, by looking at the numbers of cases and controls in each cell and calculating the probability of getting case-control values which are at least as skewed as the observed case-control values; the per-cell probabilities are combined to give an approximate log-probability of a given table. Intuitively, the lower the probability of a given case-control table arising by chance, the more likely it is to represent an underlying biological phenomenon. Given a dataset with a total of A cases and B controls, if cell i of a case-control table has a cases and b controls, we set T_i equal to the value of Fisher's exact test applied to

$$\begin{pmatrix} a & A-a \\ b & B-b \end{pmatrix}. \tag{3}$$

Thus T_i is then the two-tailed p-value for selecting a cases and b controls by chance from a total of A cases and B controls. Note that if the ith cell is empty then a and b are 0 and T_i is 1. Then, using Fisher's method [4] to combine the probabilities, the Fisher metric over the whole table is defined as

$$\sum_{i=1}^{N} -2\log(T_i) \ , \tag{4}$$

where N is the number of cells in the table. Note that, because the cells of a case-control table are not independent, the value of the Fisher metric for a table will not correspond to an exact probability; however, calculating an exact probability would be prohibitively expensive, and we hypothesize that the approximate probability used in the Fisher metric will be an effective method of scoring case-control tables.

For our comparison of metrics, we implemented Balanced Accuracy and Normalized Mutual Information, both of which are based on the risk-vs-status contingency table of true positives (TN), false negatives (FN), true negatives (TN), and false positives (FP). Balanced Accuracy, as described by Velez et al [24], is the mean of sensitivity and specificity, as shown in Eqn. 1 above, and the

corresponding metric selects the model with the highest balanced accuracy. In Normalized Mutual Information, as described by Bush et al [1], three entropies are calculated from the risk-vs-status contingency table: the row entropy, the column entropy, and a conditional entropy:

$$H(x) = -\sum_i p_i \log_2 p_i \ , \tag{5}$$

$$H(y) = -\sum_j p_j \log_2 p_j \ , \tag{6}$$

$$H(y|x) = -\sum_i p_i \sum_i \frac{p_{ij}}{p_j} \log_2 \frac{p_{ij}}{p_j} \ . \tag{7}$$

The quantity p_j is the empirical probabilities of being a case, p_i is the empirical probability of being high-risk, and p_{ij} is their joint probability. Using these entropy values, Normalized Mutual Information (NMI) is defined as:

$$NMI(y) = \frac{H(y) - H(y|x)}{H(y)}. \tag{8}$$

The Normalized Mutual Information metric selects the model with the highest value.

2.3 Numerical Analysis

We evaluated each of the four metrics over numerous different scenarios, and compared the abilities of the four metrics to distinguish a specified signal from noise (defined as a given metric assigning a higher score to the signal model than to each of the other models in the iteration). We did this by running MDR on two collections of simulated datasets: the datasets used by Velez et al [24], and a new, more comprehensive, collection of datasets generated by the GAMETES software [23,22].

In the first set of tests, we used the datasets described in Velez et al [24], which are based on a set of 70 models. Those models are based on two-way epistatic interactions with no main effect, use two minor-allele frequencies, 0.2 and 0.4, and range over the heritabilities 0.01, 0.025, 0.05, 0.1, 0.2, 0.3, and 0.4, giving a total of 14 parameter-pairs; for each of those parameter-pairs there are 5 distinct models, for a total of 70 models. Each of those models was used to generate 100 balanced datasets of 200, 400, 800, and 1600 total individuals; each dataset has 18 noise loci in addition to the 2 signal loci, for a total of 20 loci. Thus there are 7,000 datasets of each of the four sizes, for a total of 28,000 datasets. We ran MDR on each of those datasets, using each of the four metrics in turn, and evaluated how often each metric distinguished the signal model from the noise models.

In the second set of tests we tested how the ability of each metric to pick out the signal model varied with varying heritability, minor-allele frequency, and prevalence of the signal, and individual-count of the dataset overall. First, we generated 5,000 datasets for each of the heritabilities 0.01, 0.025, 0.05, 0.1, 0.2,

0.3, and 0.4; for each of those datasets, the minor-allele frequency was allowed to vary stochastically and uniformly between 0.1 and 0.5, the penetrance between 0.2 and 0.5, and the average number of individuals per cell of the case-control table between 10 and 80. Next, we generated 5,000 datasets for each of the minor-allele frequencies 0.1, 0.2, 0.3, 0.4, and 0.5; the heritabilities were allowed to vary between 0.01 and 0.4, and the penetrance and the number of individuals per cell were varied as before. Next, we generated 5,000 datasets for each of the penetrances 0.2 and 0.5, with the other parameters varying as before. Finally, we generated 5,000 datasets for each of the average individual-counts per cell 10, 20, 40, and 80, with the other parameters varying as before. We did all of these tests with a 2-locus signal, a 3-locus signal, and a 4-locus signal; in each case sufficient noise SNPs were added to achieve a total of 20 SNPs in each dataset.

The signal tables were generated using the GAMETES software[23,22]; however, there are limitations on the achievable heritabilities for high locus-counts and low minor-allele frequencies, manifested in GAMETES. Due to this limitation, for the datasets with a 3-locus signal the heritability was restricted to 0.3 or less and the minor-allele frequency to 0.2 or greater, and for the datasets with a 4-locus signal the heritability was restricted to 0.1 or less and the minor-allele frequency to 0.3 or greater.

As described above, we tested a variety of different dataset sizes. In order to make the results more comparable across different numbers of loci, we specified the dataset sizes in terms of average number of individuals per table-cell, instead of in terms of total number of individuals in the dataset, using average individual-counts per cell of 10, 20, 40, and 80 individuals. Thus, for example, we generated a 2-locus table (which has 9 cells; see Fig. 1) with a total of 90 individuals, and a 3-locus table (which has 27 cells) with a total of 270 individuals. Both tables then had an average of 10 individuals per cell. By keeping fixed the average number of individuals per cell instead of the total number of individuals in the dataset, we achieved the same degree of sparseness between the 2-locus table and the 3-locus table. If, instead, we had generated the 3-locus table with a total of just 90 individuals, its cells would be much sparser than the corresponding 2-locus table. Since the overall sparseness of individuals per cell can affect the behavior of the metrics, keeping the average number of individuals per cell constant across locus-counts makes the behavior of the metrics more comparable. Note that the goal of this approach is to maintain roughly comparable degrees of sparseness across the different levels of interaction, and we evaluate the metrics over a range of sparseness values. (With either approach to defining dataset sizes, by average-individual-count-per-cell or by total dataset size, the results will not be perfectly comparable across different numbers of interacting loci – but the central focus of this paper is to compare different metrics on similar scenarios, not across different scenarios.)

We also demonstrated the proposed new metrics by applying them to a population-based genetic study of tuberculosis (TB) that was previously analyzed using MDR by Collins et al.[2] The study analyzed 321 pulmonary TB cases and 347 healthy controls genotyped at The Bandim Health Project in Guinea Bissau.[14] Each individual was genotyped for 19 single-nucleotide

polymorphisms (SNPs) from immunological candidate genes VDR, DC-SIGN, PTX3, TLR2, TLR4, and TLR9. Collins et al imputed missing data using a frequency-based imputation and then filtered the dataset to six SNPs using ReliefF. They then applied MDR, which returned an overall best model consisting of SNPs rs2305619, rs187084, and rs1145421. In the present study, we applied all four metrics to the same filtered dataset.

Finally, we ran benchmark tests to evaluate the computation time for each metric. For each of 2-locus, 3-locus, and 4-locus interactions, we generated 10,000,000 random tables and scored each of the tables by each metric, recording the time to score each set of tables.

3 Results

3.1 Testing on the Velez Datasets

In the results of running MDR on the Velez et al[24] datasets, we group together the set of 5 models on each parameter-triple {minor-allele frequency, dataset size, heritability}, averaging the detection scores of each group. For each metric, "detection" is defined as the fraction of runs in which that metric assigned a higher score to the signal model than to each of the other models in the iteration. In every case the result was that both Variance and Fisher did as well as or better than both Normalized Mutual Information and Balanced Accuracy, with one exception: for 200-individual datasets with a minor-allele frequency of 0.2 and heritability of 0.01, Fisher had a detection score of 3.2% and Balanced Accuracy had a detection score of 3.4%. To get a high-level comparison between the various metrics, we took the overall average detection score for each metric, excluding those parameter-triples where all four metrics had detections of 0% or all four metrics had detections of 100%. By excluding the detection scores of the scenarios where either all metrics always failed or all metrics always succeeded, we concentrate on situations where the metrics differ in their effectiveness. As seen in Table 1, Variance and Fisher did about 4 to 4.5 percentage-points better than Normalized Mutual Information and Balanced Accuracy by this measure.

In order to quantify the degree to which the Variance and Fisher metrics improved over the Normalized Mutual Information and Balanced Accuracy metrics, for each parameter-triple {minor-allele frequency, dataset size, heritability} we calculated the χ^2 statistic between: the Variance metric and the Normalized

Table 1. Average detection abilities on the Velez datasets, where detection is not 0% or 100%

Metric	Detection
Variance	73.5%
Fisher	73.0%
Normalized Mutual Information	68.9%
Balanced Accuracy	69.0%

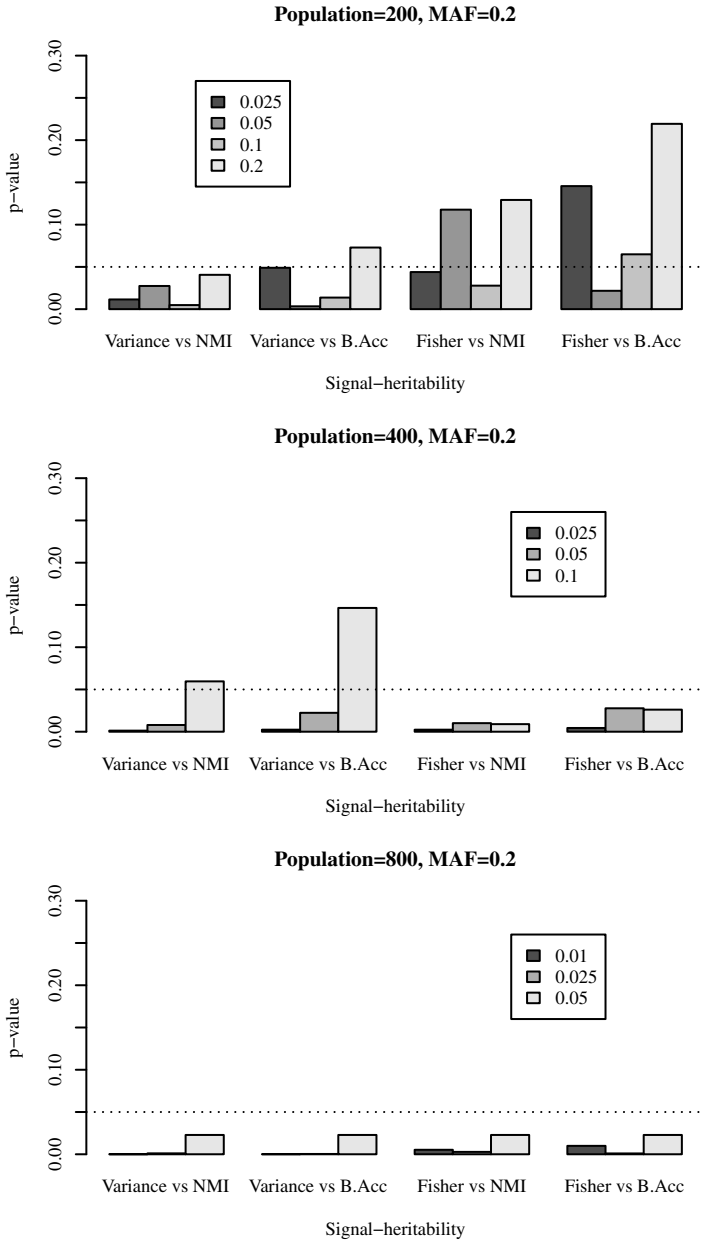

Fig. 2. Significance of χ^2 statistics for the comparisons between the Variance and Fisher metrics and the Normalized Mutual Information (NMI) and Balanced Accuracy (B.Acc) metrics, for the signal-heritabilities listed in the legends

Mutual Information metric; Variance and Balanced Accuracy; Fisher and Normalized Mutual Information; and Fisher and Balanced Accuracy. For example, for the χ^2 statistic between the Variance metric and the Normalized Mutual Information metric for a given scenario, we constructed a table of the success and failure counts for each of the metrics under that scenario, and calculated the R chisq.test function on that table. We then calculated the statistical significance of each χ^2 statistic. We show selected results as follows; in the results not shown, differences between the metrics were usually marginally significant or not significant. As mentioned above, in all scenarios where there was a significant difference, Variance and Fisher scored better than the Normalized Mutual Information and Balanced Accuracy metrics.

For 200-individual datasets and a minor-allele frequency of 0.2, the improvement of the Variance metric over the Normalized Mutual Information metric was significant at the 0.05 level or better for heritabilities of 0.025, 0.05, 0.1, and 0.2 (Fig. 2). The improvement of the Variance metric over the Balanced Accuracy metric was very significant for heritabilities of 0.05, and 0.1 and marginally signficant for heritabilities of 0.025 and 0.2. The comparison between the Fisher metric and the Normalized Mutual Information and Balanced Accuracy metrics was mixed, as seen in the figure.

For 400-individual datasets and a minor-allele frequency of 0.2, the improvement of the Variance and Fisher metrics over the Normalized Mutual Information and Balanced Accuracy metrics was very significant for heritabilities of 0.025, 0.05, and 0.1, except for the Variance metric in the scenario with heritability of 0.1.

For 800-individual datasets and a minor-allele frequency of 0.2, the improvement of the Variance and Fisher metrics over the Normalized Mutual Information and Balanced Accuracy metrics was very significant for heritabilities of 0.01, 0.025, and 0.05.

3.2 Comprehensive Testing on GAMETES Datasets

In the second part of the study we tested the effectiveness of each metric over a wide range of scenarios; we found that the two new metrics, Variance and Fisher, always did as well as or better than Normalized Mutual Information and Balanced Accuracy, with one exception: in the scenario of a 2-way signal with heritability of 0.4, the Normalized Mutual Information and Balanced Accuracy metrics score 99.98% while the Variance metric scored 99.96%; in that scenario the Fisher metric scored 100%. The only scenarios where the two older metrics were close to the two new metrics were scenarios where all four metrics had scores near 100%, usually because the heritability of the signal was high, making it easy to detect. Whenever the metrics scored less than 85%, the two new metrics outscored the two older metrics by at least two percentage points. Thus we see that when the signal is relatively easy to find the two new metrics do as well as or better than the older metrics, and when the signal is harder to detect the new metrics do significantly better – by ten percentage points or more in five of the scenarios.

These observations are made more precise by using a χ^2 analysis. As in the analysis of the Velez datasets, we calculated the χ^2 statistic between: the

Variance metric and the Normalized Mutual Information metric; Variance and Balanced Accuracy; Fisher and Normalized Mutual Information; and Fisher and Balanced Accuracy. We then calculated the statistical significance of each χ^2 statistic. In most scenarios, the improvement of the new metrics over the older ones is highly significant. We found three categories of performance: When all four metrics detect the signal correctly 99% of the time or more, the χ^2 comparisons between the metrics showed no significant difference, with p-values of 0.3 or greater. When the metrics had a correct detection level between 85% and 99%, the χ^2 comparisons showed the new metrics significantly better than the older metrics, with p-values between 0.04 and 0.001. And when the signal was harder to detect, with the metrics finding the correct signal less than 85% of the time, the improvement of the new metrics over the older ones was highly significant, with p-values less than 0.001.

In the comprehensive testing using GAMETES datasets there were 5,000 datasets for each scenario, as compared with the 500 datasets per scenario in the Velez datasets; we see that with the greater resolution afforded by the larger number of datasets, the improvement inherent in the new metrics becomes crystal clear.

3.3 Demonstration Data and Benchmarks

We also tested the four metrics on a tuberculosis dataset that had previously been evaluated using MDR, which found an overall best model consisting of SNPs rs2305619, rs187084, and rs1145421. In our tests, all four of the metrics identified that model as best overall.

The computation times for the Variance, Normalized Mutual Information, and Balanced Accuracy metrics are similar (Tbl. 2). The Fisher metric, being more computationally intensive, takes considerably longer to run; however, that run-time could be improved dramatically by caching the probability calculations.

Table 2. Time in seconds to calculate each metric on 10,000,000 tables for 2-locus to 4-locus interactions, running in Java on a 2.26 GHz Intel Xeon with single-threading

Metric	2-locus	3-locus	4-locus
Variance	2.81	6.87	15.84
Fisher	138.99	248.56	593.33
Normalized Mutual Information	5.6	8.08	13.95
Balanced Accuracy	2.33	6.84	21.14

4 Discussion

The ability to discover the connections between genotype and phenotype is central to genomics research, but it continues to be challenging. It was over a decade ago that Risch and Merikangas first seriously proposed the testing of all known

SNPs in the human genome for disease association either directly or by LD with other SNPs [16]. Today, it is becoming cost effective to measure a million SNPs with widely-available human SNP arrays, but the tools used to analyze that data need to improve as well [12]. Part of that improvement is to ensure that all of the information contained within each dataset is fully employed.

The reduced-dimensionality high-risk/low-risk contingency table produced by MDR contains less information than the case-control table for the model it represents – each cell of the case-control table contains information about the numbers of cases and controls in that cell, and that information is omitted when the cases and controls are summed into the contingency table. Thus, by defining our metrics, for the purpose of model-selection only, directly on the case-control tables instead of on the contingency tables, we are able to make better use of that information in selecting a model. Once selected, the model is reduced to a high-risk/low-risk contingency table in the usual way.

We find that the Variance and Fisher metrics do as well as or significantly better than Normalized Mutual Information and Balanced Accuracy in all of the wide variety scenarios in which they were tested, as measured in terms of detection ability. The improvement is especially strong when the signal is difficult to detect, which is exactly the scenario where improvement is most desirable. The Fisher metric is of particular value because it gives a direct measure of how unlikely a given model is to have arisen by chance, and therefore of how likely the model is to reflect an underlying biological phenomenon. However, it takes substantially more computation time than any of the other metrics tested. Given that the Variance metric closely parallels the Fisher metric in all regimes tested, we recommend the Variance metric for use with MDR going forward.

Acknowledgments. This work was supported by NIH grants LM009012, LM010098, and AI59694.

References

1. Bush, W.S., Edwards, T., Dudek, S., McKinney, B., Ritchie, M.: Alternative contingency table measures improve the power and detection of multifactor dimensionality reduction. BMC Bioinformatics 9, 238 (2008)
2. Collins, R.L., Hu, T., Wejse, C., Sirugo, G., Williams, S., Moore, J.: Multifactor dimensionality reduction reveals a three-locus epistatic interaction associated with susceptibility to pulmonary tuberculosis (2012) (manuscript submitted for publication)
3. Cordell, H.: Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum. Mol. Genet. 11, 2463–2468 (2002)
4. Fisher, R.: Statistical methods for research workers. Genesis Publishing Pvt. Ltd. (1925)
5. Hahn, L., Moore, J.: Ideal discrimination of discrete clinical endpoints using multilocus genotypes. In Silico Biol. 4, 0016 (2004)
6. Hahn, L., Ritchie, M., Moore, J.: Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. Bioinformatics 19, 376–382 (2003)

7. Moore, J.H.: Computational analysis of gene-gene interactions in common human diseases using multifactor dimensionality reduction. Expert Rev. Mol. Diagn. 4, 795–803 (2004)
8. Moore, J.: The ubiquitous nature of epistasis in determining susceptibility to common human diseases. Hum. Hered. 56, 73–82 (2003)
9. Moore, J.: A global view of epistasis. Nat. Genet. 37, 13–14 (2005)
10. Moore, J.: Genome-wide analysis of epistasis using multifactor dimensionality reduction: feature selection and construction in the domain of human genetics. In: Zhu, X., Davidson, I. (eds.) Knowledge Discovery and Data Mining: Challenges and Realities with Real World Data, pp. 17–30. IGI Press, Hershey (2007)
11. Moore, J., Gilbert, J., Tsai, C., Chiang, F., Holden, W., Barney, N., White, B.: A flexible computational framework for detecting, characterizing, and interpreting statistical patterns of epistasis in genetic studies of human disease susceptibility. J. Theor. Biol. 241, 252–261 (2006)
12. Moore, J., Williams, S.: New strategies for identifying gene-gene interactions in hypertension. Ann. Med. 34, 88–95 (2002)
13. Moore, J., Williams, S.: Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. BioEssays 27, 637–646 (2005)
14. Olesen, R., Wejse, C., Velez, D., Bisseye, C., Sodemann, M., Aaby, P., Rabna, P., Worwui, A., Chapman, H., Diatta, M., Adegbola, R., Hill, P., Stergaard, L., Williams, S., Sirugo, G.: Dc-sign (cd209), pentraxin 3 and vitamin d receptor gene variants associate with pulmonary tuberculosis risk in West Africans. Genes and Immunity 8(suppl. 6), 456–467 (2007)
15. Rea, T., Brown, C., Sing, C.: Complex adaptive system models and the genetic analysis of plasma hdl-cholesterol concentration. Perspect. Biol. Med. 49, 490–503 (2006)
16. Risch, N., Merikangas, K.: The future of genetic studies of complex human disease. Science 273, 1516–1517 (1996)
17. Ritchie, M., Hahn, L., Moore, J.: Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet. Epidemiol. 24, 150–157 (2003)
18. Ritchie, M., Hahn, L., Roodi, N., Bailey, L., Dupont, W., Parl, F., Moore, J.: Multifactor dimensionality reduction reveals high-order interactions among estrogen metabolism genes in sporadic breast cancer. Am. J. Hum. Genet. 69, 138–147 (2001)
19. Sing, C., Stengard, J., Kardia, S.: Genes, environment, and cardiovascular disease. Arterioscler. Thromb. Vasc. Biol. 23, 1190–1196 (2003)
20. Templeton, A.: Epistasis and complex traits. In: Wade, M., Brodie III, B., Wolf, J. (eds.) Epistasis and Evolutionary Process. Oxford University Press, New York (2000)
21. Thornton-Wells, T., Moore, J., Haines, J.: Genetics, statistics, and human disease: analytical retooling for complexity. Trends Genet. 20, 640–647 (2004)
22. Urbanowicz, R., Kiralis, J., Fisher, J., Moore, J.: Predicting the difficulty of pure, strict, epistatic models: metrics for simulated model selection. BioData Mining 5(1), 15 (2012)
23. Urbanowicz, R., Kiralis, J., Sinnott-Armstrong, N., Heberling, T., Fisher, J., Moore, J.: Gametes: a fast, direct algorithm for generating pure, strict, epistatic models with random architectures. BioData Mining 5(1), 16 (2012)
24. Velez, D., White, B., Motsinger, A., Bush, W., Ritchie, M., Williams, S., Moore, J.: A balanced accuracy metric for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. Genet. Epidemiol. 31, 306–315 (2007)